Assignment 2

Participate in the *Digit Recognizer* competition at *kaggle.com*.

Link: https://www.kaggle.com/c/digit-recognizer

Follow the below steps to submit your first prediction:

1. Create a user id and password in kaggle.com
2. Join the *Digit Recognizer* competition
3. Download the train and test data
4. Merge the two datasets
5. Use *statsbuddy.net* & *R Studio Cloud* to predict the *label* column in test data
6. Manually delete all columns, other than *Id* and *label*, from the prediction file
7. Submit your prediction on *kaggle.com* and get your score and rank

Try to improve your score and rank by trying:

- Different sets of independent variables, depending on their level of influence
- Different machine learning models (e.g., Decision Tree, and Random Forest, SVM, XGBoost, etc.)
- Different hyper-parameters (e.g., number of levels, and number of trees, as applicable)
- Principal Component Analysis for dimensionality reduction
- Ensemble Methods

Deliverables:

Write a two-page report describing your approach and journey. Describe what you tried (and what you didn't try) and why (very important!). No R codes are required. Submit the report and a screenshot of the leaderboard from *kaggle.com* that shows your score and rank.

Distribution of Points:

10 points for your score and rank, and 10 points for the analysis and report. The first 10 points will be distributed based on your relative rank in the class (i.e., 10 for the top 10% students in terms of rank, 9 for the next 20% students, 8 for the next 30% students and 7 for the remaining students.)